

Nonlinear Contraction Tools for Constrained Optimization

Jonathan Soto and Jean-Jacques E. Slotine

Abstract

This paper describes new results linking constrained optimization theory and nonlinear contraction analysis. Generalizations of Lagrange parameters are derived based on projecting system dynamics on the tangent space of possibly time-varying constraints. The paper formalizes the intuition that, just as convexity rather than linearity is the key property in optimization, contraction rather than linearity is the key dynamical property in this context.

Index Terms

Nonlinear dynamics, nonlinear control, constrained optimization, Lagrange parameters, sliding control, contraction analysis

I. INTRODUCTION

After presenting some initial examples that show the relations between contraction and optimization, we derive a contraction theorem for nonlinear systems with equality constraints. The method is applied to examples in differential geometry and biological systems. A new physical interpretation of Lagrange parameters is provided. In the autonomous case, we derive a new algorithm to solve minimization problems. Next, we state a contraction theorem for nonlinear systems with inequality constraints. Finally, we state another contraction theorem for nonlinear systems with time-varying equality constraints. A new generalization of time varying Lagrange parameters is given. In the autonomous case, we provide a solution for a new class of optimization problems, minimization with time-varying constraints.

In the following we consider an n -dimensional time-varying system of the form:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), t) \quad (1)$$

where $\mathbf{x} \in R^n$ and $t \in R^+$ and \mathbf{f} is $n \times 1$ nonlinear vector function which is assumed to be real and smooth in the sense that all required derivatives exist and are continuous.

II. PRELIMINARIES

A. On contraction analysis

Our results are based on nonlinear contraction theory [1], a viewpoint on incremental stability whose basic result we now recall. Historically, ideas closely related to contraction can be traced back to [24] and even to [25] (see also [26], [27], and e.g. [28] for a more exhaustive list of related references). As pointed out in [1], contraction is preserved through a large variety of systems combinations, and in particular it represents a natural tool for the study and design of nonlinear state observers, and by extension, of synchronization mechanisms [29].

The basic result of contraction analysis can be stated as (for more details we refer to [1]):

Theorem 1. *Denote by $\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$ the Jacobian matrix of \mathbf{f} with respect to \mathbf{x} . Assume that there exists a complex square matrix $\Theta(\mathbf{x}, \mathbf{t})$ such that the Hermitian matrix $\Theta(\mathbf{x}, \mathbf{t})^* \Theta(\mathbf{x}, \mathbf{t})$ is uniformly positive definite, and the Hermitian part \mathbf{F}_H of the matrix*

$$\mathbf{F} = \left(\dot{\Theta} + \Theta \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right) \Theta^{-1}$$

is uniformly negative definite. Then, all system trajectories converge exponentially to a single trajectory, with convergence rate $|\sup_{\mathbf{x}, \mathbf{t}} \lambda_{\max}(\mathbf{F}_H)| > 0$. The system is said to be contracting, \mathbf{F} is called its generalized Jacobian, and $\Theta(\mathbf{x}, \mathbf{t})^ \Theta(\mathbf{x}, \mathbf{t})$ its contraction metric. The contraction rate is the absolute value of the largest eigenvalue (closest to zero, although still negative) $\lambda = |\lambda_{\max}(\mathbf{F}_H)|$.*

B. Optimization theory

Optimization means finding "best available" values of some objective function given a defined domain, including a variety of different types of objective functions and different types of domains. We give the sufficient conditions for the point \mathbf{x}^* to be a minimum of a minimization problem. There are different classes of problems which are stated below. (for more details we refer to [2])

In unconstrained optimization we search for the minimum over the whole range of variables. The generic form is:

$$\min U(\mathbf{x}) \quad (2)$$

The second-order sufficient conditions in the unconstrained case are given by theorem 2.

Theorem 2. *Let $U \in C^2(R^n, R)$ be a function defined on a region in which the point \mathbf{x}^* is an interior point. Suppose in addition that*

$$\nabla U(\mathbf{x}^*) = 0 \quad \text{and} \quad \nabla^2 U(\mathbf{x}^*) \text{ is positive definite}$$

\mathbf{x}^ is a solution of problem 2.*

In constrained optimization, the Karush-Kuhn-Tucker conditions (also known as the Kuhn-Tucker or KKT conditions) are sufficient for a solution in nonlinear programming to be optimal, provided that some regularity conditions are satisfied. The generic form is with m constraints is:

$$\min U(\mathbf{x}) \quad \text{subject to (s.t)} \quad \mathbf{h}(\mathbf{x}) = \mathbf{0} \quad (3)$$

The second-order sufficient conditions in the constrained case are given by theorem 3.

Theorem 3. *Let $U(\mathbf{x}) \in C^2(R^n, R)$ and $\mathbf{h}(\mathbf{x}) \in C^2(R^n, R^m)$ be smooth functions and \mathbf{x}^* is a regular point (the columns of $\nabla \mathbf{h}(\mathbf{x}^*)$ are linearly independent). Suppose in addition that*

$$\nabla L(\mathbf{x}^*) = 0 \quad \text{and} \quad \mathbf{y}' \nabla^2 L(\mathbf{x}^*) \mathbf{y} > 0$$

where $\mathbf{y} \nabla \mathbf{h}(\mathbf{x}^) = 0$, λ are the Lagrange parameters and L is the lagrangian function defined as $L = U + \lambda \mathbf{h}$.*

\mathbf{x}^ is a solution of problem 3.*

C. Examples

We present some cases - unconstrained optimization, duality theory - in which both theories are linked. The following example in [3] is the starting point of this research.

Example II.1.: Consider a gradient autonomous system $\dot{\mathbf{x}} = -\nabla U(\mathbf{x})$, contracting in an identity metric, $\Theta = I$. As it is autonomous and contracting in a time independent metric, it has a unique equilibrium point because

$$\frac{d}{dt}(\nabla U) = \mathbf{F}(\nabla U)$$

which implies exponential convergence of $\dot{\mathbf{x}}$ to zero. \mathbf{x} converges to a constant, \mathbf{x}^* . This point has zero speed, $\nabla U(\mathbf{x}^*) = 0$. The condition of contraction at this equilibrium point is the positive definiteness of $\nabla^2 U(\mathbf{x}^*)$. These are the two sufficient conditions (theorem 2) to prove that \mathbf{x}^* is a solution to problem 2

Contraction theory is also related to differential geometry through example II.2. [19]

Example II.2.: The condition of contraction of the system 1 for a positive definite metric $\mathbf{g} = \Theta' \Theta$ that verifies a parallel propagation is that the hermitian part of

$$\mathbf{F} = \Theta D\mathbf{f}(\mathbf{x}, t) \Theta^{-1} = \Theta \left(\frac{\partial \mathbf{f}(\mathbf{x}, t)}{\partial \mathbf{x}} + \sum_i \Gamma_{ij}^k \mathbf{f}^i(\mathbf{x}, t) \right) \Theta^{-1}$$

has a uniformly negative maximum eigenvalue. D is the covariant derivative and the Christoffel term is defined as $\sum_k \Gamma_{ij}^k g_{kl} = \frac{1}{2} \left(\frac{\partial g_{il}}{\partial q^j} + \frac{\partial g_{jl}}{\partial q^i} - \frac{\partial g_{ij}}{\partial q^l} \right)$.

The idea is to use a parallel propagation of the tensor Θ , [19].

$$\dot{\Theta} = \Theta \sum_i \Gamma_{ij}^k \dot{\mathbf{x}}^i$$

The geometric interpretation of the covariant derivative is the projection of the directional derivative [7] on that submanifold, thus the tangential part of the directional derivative. It can be written as:

$$D_X Y = (d_X Y)^{tang} = d_X Y - \langle d_X Y, \nu \rangle \nu$$

where $d_X Y$ is the directional derivative along the X direction and ν is the normal vector to the submanifold.

When writing $\delta \dot{z} = \Theta D\mathbf{f} \Theta^{-1} \delta z$, it means that the virtual speed is constrained to belong to 'some' manifold. This idea of projection is fundamental to prove contraction for constrained dynamical systems.

With example II.1, we show that contraction and unconstrained optimization are linked through gradient systems. Another important domain in optimization is duality. It is very useful to use the dual formulation instead of the primal for a variety of reasons. Sometimes the dual problem has a closed form solution or the algorithm to find the minimum is much faster. In example II.3, we show that contraction and duality are related using gradient systems.

The Legendre transformation is defined as follows for $y \in R^n$:

$$U^*(\mathbf{y}) = \sup_{\mathbf{x} \in R^n} (\mathbf{x}' \mathbf{y} - U(\mathbf{x})) \quad (4)$$

The conjugate function U^* is a convex function since it is the pointwise supremum of linear functions.

Example II.3.: The system $\dot{\mathbf{x}} = -\nabla U(\mathbf{x})$ is contracting for the identity metric and also smooth (existence of derivatives) if and only if the system $\dot{\mathbf{y}} = -\nabla U^*(\mathbf{y})$ is contracting for the identity metric and also smooth.

To prove the if part we use the fact that strict convexity and smoothness are dual properties, [23].

For the only if part, we apply the first part of the proof to the function U^* . We use the fact that as U^* convex, close (the domain of U is closed) and proper (it never takes the value $-\infty$ and the set $\text{dom}_g = [x|g(x) < \infty]$ is nonempty), $U^{**} = U$

In [1] combination properties of contracting systems are of great importance. In example II.4 we derive a new result that cannot be achieved with classical minimization theory.

Example II.4.: If there are two unconstrained minimizations which have respectively their minimum at x_1^* and x_2^* , the sum of both cost functions may not have a minimum. Using only classical minimizations theorems we cannot conclude anything when summing both systems.

If the gradient systems issued from the two precedent minimization problems are contracting, the sum of the gradient systems, is still contracting. Therefore the sum of the two cost functions has a minimum

In the following all the points are regular. This means that $\nabla \mathbf{h}(\mathbf{x})\nabla \mathbf{h}'(\mathbf{x})$ has full rank and hence is invertible. That is the *only* hypothesis needed in the following analysis.

III. CONTRACTION THEORY WITH EQUALITY CONSTRAINTS

A. Starting on the constraints: contraction theory

We define the sets $S = [\mathbf{x} / \mathbf{h}(\mathbf{x}) = 0]$ and $M(\mathbf{x}) = [y / y\nabla \mathbf{h}(\mathbf{x}) = 0]$. S is the space of the constraints. M is the tangent subspace of \mathbf{h} at point \mathbf{x} , it is a k -submanifold. (a k submanifold is defined as $\nabla \mathbf{h}(\mathbf{x})$ having full rank using the implicit function theorem).

We define the operator [4]

$$P(\mathbf{x}) = 1 - \nabla \mathbf{h}'(\mathbf{x})[\nabla \mathbf{h}(\mathbf{x})\nabla \mathbf{h}(\mathbf{x})']^{-1}\nabla \mathbf{h}(\mathbf{x})$$

It is a symmetric orthogonal projection operator onto $M(\mathbf{x})$ because it verifies $P(\mathbf{x})P(\mathbf{x}) = P(\mathbf{x})$ and $P(\mathbf{x})y \in M(\mathbf{x})$.

1) *Contraction theorem :*

Theorem 4. *The condition of contraction of the constrained dynamic system*

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) \text{ s.t. } \mathbf{h}(\mathbf{x}) = 0$$

is that there exists a metric, Θ , uniformly positive definite and that the hermitian part of :

$$\mathbf{F} = \left(\dot{\Theta}(\mathbf{x}, t) + \Theta(\mathbf{x}, t) P(\mathbf{x}) \left(\frac{\partial \mathbf{f}}{\partial \mathbf{x}} + \lambda(\mathbf{x}, t) \frac{\partial^2 \mathbf{h}}{\partial \mathbf{x}^2} \right) \right) \Theta(\mathbf{x}, t)^{-1} \text{ on } S(\mathbf{x})$$

has a maximum eigenvalue uniformly negative. $P(\mathbf{x})$ is the projection operator onto the tangent space $M(\mathbf{x})$ and $\lambda(\mathbf{x}, t) = -\mathbf{f}(\mathbf{x}, t) \nabla \mathbf{h}' [\nabla \mathbf{h} \nabla \mathbf{h}']^{-1}$. The initial condition must verify the constraint.

The proof has three parts. We project the system by adding a term related to the Lagrange parameters. We compute the condition of contraction of this new system. Finally, we prove contraction behavior.

The initial problem is

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) \text{ s.t. } \mathbf{h}(\mathbf{x}) = 0$$

Physically \mathbf{f} is the velocity of the system going away from the constraints. To verify the constraints, the systems's velocity has to be tangent to the constraints at \mathbf{x} . We project the system onto $M(\mathbf{x})$ using the projector $P(\mathbf{x})$.

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) + \lambda(\mathbf{x}, t) \nabla \mathbf{h}(\mathbf{x})$$

We note $\lambda(\mathbf{x}, t) = -\mathbf{f}(\mathbf{x}, t) \nabla \mathbf{h}'(\mathbf{x}) [\nabla \mathbf{h}(\mathbf{x}) \nabla \mathbf{h}(\mathbf{x})']^{-1}$. These are not exactly the Lagrange parameters. But if $\mathbf{f} = \nabla U(\mathbf{x})$, we recognize the Lagrange parameters. This is a first generalization of the Lagrange parameters. The new dynamical system is $\dot{\mathbf{x}} = \nabla U(\mathbf{x}) + \lambda(\mathbf{x}) \nabla \mathbf{h}(\mathbf{x}) = \nabla L(\mathbf{x})$ where L is the usual Lagrangian function. In the following the projected system will be known as Lagrangian dynamics. This result gives a new insight about Lagrange parameters. The term $\lambda(\mathbf{x}, t) \nabla \mathbf{h}(\mathbf{x})$ is a reaction force allowing the system to stay on the constraints. A Lagrange parameter is the scalar value that gives the magnitude of the force along the orthogonal direction to a particular constraint in order to have a tangential speed. Lagrange parameters as reaction forces has already been investigated using the Lagrange equation [6] [5]. Anytime, the constraints are verified because the system starts on the constraints and the velocity is tangential to the constraints. If initially the constraints were not verified, this property would not be true.

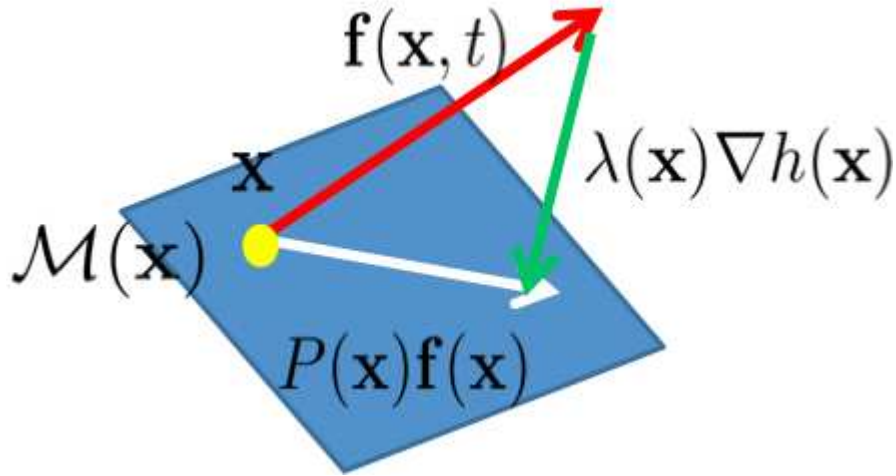


Fig. 1. Projected speed of the dynamic system

We consider a virtual displacement $\delta \mathbf{x}$ and a positive definite metric, $\Theta(\mathbf{x}, t)$. The function $\delta \mathbf{x}' \Theta(\mathbf{x}, t)' \Theta(\mathbf{x}, t) \delta \mathbf{x}$ is the distance associated to the first fundamental form of the k -submanifold $M(\mathbf{x})$. We define $\delta \mathbf{z} = \Theta(\mathbf{x}, t) \delta \mathbf{x}$.

We compute $\frac{d}{dt}(\delta \mathbf{z}' \delta \mathbf{z}) = 2\delta \mathbf{z}' \delta \dot{\mathbf{z}}$. We calculate $\delta \dot{\mathbf{z}} = \dot{\Theta}(\mathbf{x}, t) \delta \mathbf{x} + \Theta(\mathbf{x}, t) \delta \dot{\mathbf{x}}$ where $\delta \dot{\mathbf{x}} = \left(\frac{\partial f}{\partial \mathbf{x}} + \lambda(\mathbf{x}, t) \frac{\partial^2 \mathbf{h}}{\partial \mathbf{x}^2} + \frac{\partial \lambda(\mathbf{x}, t)}{\partial \mathbf{x}} \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \right) \delta \mathbf{x}$

We compute a long calculation $\frac{\partial \lambda(\mathbf{x}, t)}{\partial \mathbf{x}}$. This is done in [4] and [8].

$$\delta \dot{\mathbf{x}} = P(\mathbf{x}) \left(\frac{d\mathbf{f}}{d\mathbf{x}} + \lambda(\mathbf{x}, t) \frac{\partial^2 \mathbf{h}}{\partial \mathbf{x}^2} \right) \delta \mathbf{x} - \nabla \mathbf{h}(\mathbf{x})' [\nabla \mathbf{h}(\mathbf{x}) \nabla \mathbf{h}(\mathbf{x})']^{-1} \delta \mathbf{x} \frac{\partial^2 \mathbf{h}}{\partial \mathbf{x}^2} \dot{\mathbf{x}} \quad (5)$$

The first term belongs to $M(\mathbf{x})$, the second term to $M(\mathbf{x})^\perp$. The equation can be rewritten :

$$\delta \dot{\mathbf{x}} = \delta \dot{\mathbf{x}}^\parallel + \delta \dot{\mathbf{x}}^\perp$$

In order to have contraction behavior the first term of $\delta \dot{\mathbf{x}}^\parallel$ must be uniformly bounded. Finally

$$\frac{d}{dt}(\delta \mathbf{z}' \delta \mathbf{z}) = 2\delta \mathbf{z}' \left(\dot{\Theta}(\mathbf{x}, t) + \Theta(\mathbf{x}, t) P(\mathbf{x}) \left(\frac{\partial f}{\partial \mathbf{x}} + \lambda(\mathbf{x}, t) \frac{\partial^2 \mathbf{h}}{\partial \mathbf{x}^2} \right) \right) \Theta(\mathbf{x}, t)^{-1} \delta \mathbf{z}$$

Example III.1.: The major goal of control theory is to find the control input that will make the system behave in a defined way. In mathematical terms, find $\hat{\mathbf{u}}$ such that

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) + \mathbf{u}(\mathbf{x}, t)$$

has a specific behavior. In the case of projected contraction theory identifying $\hat{\mathbf{u}}(\mathbf{x}, t) = \lambda \nabla \mathbf{h}$, we find a control that makes the system evolve on the constraint.

The following example shows how projected contraction can be applied to general dynamic systems (not only to gradient systems, i.e. minimization problems).

Example III.2.: The system 6 arises from the study of a mathematical model for the spread of an infectious disease in a population with a fixed total size. The variables s , e , i , and r represent fractions of the population that are susceptible, exposed (in the latent period), infectious, and recovered, respectively. All parameters are assumed to be nonnegative, we assume also $\epsilon > 0$ and $\gamma > 0$.

$$\begin{aligned}\dot{s} &= b - bs - \lambda is + \alpha is + \delta r \\ \dot{e} &= \lambda is - (\epsilon + b)e + \alpha ie \\ \dot{i} &= \epsilon e - (\gamma + \alpha + b)i + \alpha i^2 \\ \dot{r} &= \gamma i - (b + \delta)r + \alpha ir\end{aligned}\tag{6}$$

The biological feasible region is the following invariant simplex $s + e + i + r = 1$, the sum of the different populations is the total population. We project the dynamic system. The value of $\lambda = -f\nabla h'[\nabla h\nabla h']^{-1} = -\frac{1}{4}(b - bs + \alpha is - be + \alpha ie - (\alpha + b)i + \alpha i^2 - br + \alpha ir)$. The jacobian of the system is:

$$\frac{df}{dx} = \begin{pmatrix} -b - \lambda i + \alpha i & 0 & (-\lambda + \alpha)s & \delta \\ \lambda i & -(\epsilon + b) + \alpha i & \lambda s + \alpha i & 0 \\ 0 & \epsilon & -(\gamma + \alpha + b) + \alpha i & 0 \\ 0 & 0 & \alpha e & -(b + \delta) + \alpha i \end{pmatrix}$$

The projection matrix is :

$$P = 1 - \nabla h'(\mathbf{x})[\nabla h(\mathbf{x})\nabla h(\mathbf{x})']^{-1}\nabla h(\mathbf{x}) = \frac{1}{4} \begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix}$$

On figure 2 , the system converges to $\mathbf{x}^* = [0.050, 0.542, 0.193, 0.213]$. This equilibrium point corresponds to an endemic population. The condition of contraction has to be calculated for a metric, $\Theta(\mathbf{x})$. In this case the metric is the complex matrix that diagonalizes the jacobian times the projection matrix.

$$F = \Theta(\mathbf{x})P\frac{df}{dx}\Theta(\mathbf{x})^{-1}$$

We do not have $\lambda \nabla^2 \mathbf{h}$ in the contraction condition because the constraint is linear. There is just one eigenvalue that is equal to zero. We get rid of it by using a three times four projection matrix. The vectors of this matrix are a basis for M . Using the interlacing theorem, the eigenvalues of the projected matrix are all negative, the system is contracting.

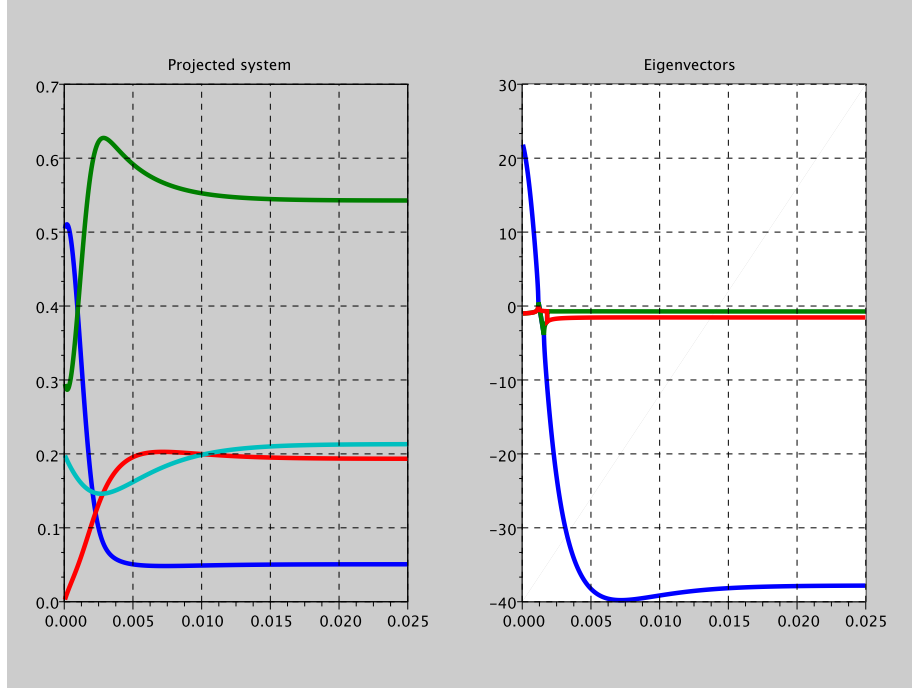


Fig. 2. Projected speed of the dynamic system

2) *An algorithm for solving minimization problems:* As shown in example II.1 an autonomous contracting dynamic system has a unique equilibrium point \mathbf{x}^* .

$$\dot{\mathbf{x}} = -\nabla U(\mathbf{x}) \text{ s.t. } \mathbf{h}(\mathbf{x}) = 0 \quad (7)$$

The projected system, which we call Lagrangian dynamic, is $\dot{\mathbf{x}} = -\nabla L(\mathbf{x})$. It verifies $\nabla L(\mathbf{x}^*) = 0$. The dynamic system also satisfies the constraint $\mathbf{h}(\mathbf{x}^*) = 0$ because we start on the constraint and the speed of the system is always tangential to the constraints. These two conditions ensure us that at anytime $\mathbf{h}(\mathbf{x}) = 0$.

The system being autonomous, the condition of contraction is, 4

$$\ddot{\mathbf{x}} = -P(\mathbf{x})\nabla^2 L(\mathbf{x})\dot{\mathbf{x}}$$

As $\dot{\mathbf{x}} \in M(\mathbf{x})$, we have : $\dot{\mathbf{x}} = P(\mathbf{x})\dot{\mathbf{x}}$. Thus: $\ddot{\mathbf{x}} = -P(\mathbf{x})\nabla^2 L(\mathbf{x})P(\mathbf{x})\dot{\mathbf{x}}$. The condition of contraction is

$$\mathbf{x}^{*'} P'(\mathbf{x}^*) \left(\frac{d^2 U}{d\mathbf{x}^2} + \lambda(\mathbf{x}^*) \frac{\partial^2 \mathbf{h}}{\partial \mathbf{x}^2} \right) P(\mathbf{x}^*) \mathbf{x}^* = \mathbf{y}' \nabla^2 L(\mathbf{x}^*) \mathbf{y} > \eta > 0$$

with $\mathbf{y} = \mathbf{x}P(\mathbf{x}^*) \in M(\mathbf{x}^*)$ by definition of $P(\mathbf{x}^*)$.

These are the three conditions for the existence of a minimum to problem 3. This is a new relationship between minimization and dynamical systems using contraction theory. This has been investigated [9].

This idea of projection creates a new dynamical system. It is a new way of solving minimization problems. We give some examples. The first one is done in great detail.

Example III.3.: The minimization of the length of a square in a circle is

$$\min x + y \text{ s.t. } x^2 + y^2 = 1$$

The gradient dynamical associated is $\dot{x} = -\nabla f = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$ s.t. $x^2 + y^2 = 1$. After projection

$$\dot{x} = \begin{pmatrix} -1 + x^2 + xy \\ -1 + y^2 + xy \end{pmatrix}$$

where $\lambda = -\frac{\nabla f \nabla h'}{\nabla h \nabla h'} = -\frac{x+y}{2}$. The eigenvalue of F is $x + y$. The system is contracting in $x + y \leq 0$. As the system approaches the solution, $\lambda \rightarrow \lambda^*$. Figure 3 shows that the dynamic system always stays on the constraint while it goes to the minimum.

Example III.4.: Consider the minimum distance from a point to a sphere problem

$$\min \frac{1}{2}(x^2 + y^2 + (z - 2)^2) \text{ s.t. } h(x) = x^2 + y^2 + z^2 - 1 = 0$$

The evident solution of this problem is the projection of the point $(0, 0, 2)$ on the sphere: $(0, 0, 1)$. The projected system is:

$$\dot{x} = \nabla U + \lambda \nabla h = \begin{pmatrix} x + \lambda x \\ y + \lambda y \\ z - 2 + \lambda z \end{pmatrix}$$

with $\lambda = 2z - 1$.

On figure 4, the system evolves on the constraint. it converges towards the expected solution.

Example III.5.: The minimum distance from a point to an ellipse is chosen to show that the lack of symmetry does not prevent convergence of the algorithm. We search the distance from a point to the ellipse.

$$\min \frac{1}{2} \|\mathbf{x} - \mathbf{x}_1\|^2 \text{ s.t. } h(\mathbf{x}) = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1 = 0$$

Consider $a = 15; b = 5; c = 3$; and the absolute minimum being 1, 4, 2. On figure 5 the solution corresponds to the projection of the absolute minimum on the ellipsoid. (The red dotted line just indicates where the projection point is). As usual the convergence's speed is very fast compared to matlab.

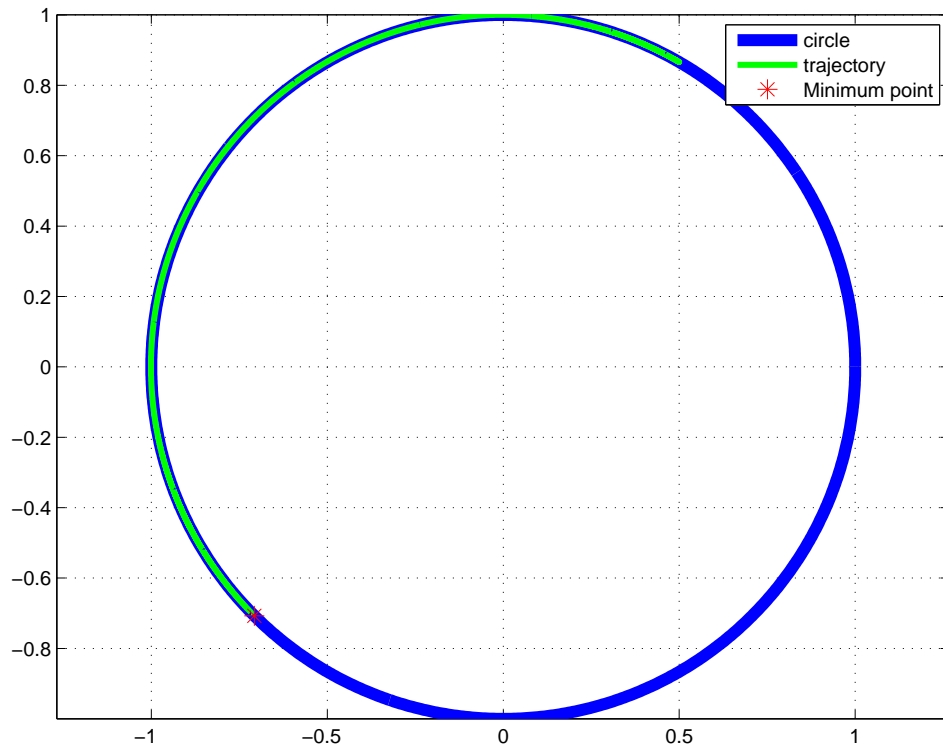


Fig. 3. Minimization of the length of a square in a circle in 2D

Example III.6.: The minimization problem with two constraints

$$\min \frac{1}{2} \|\mathbf{x} - \mathbf{x}_1\|^2 \quad s.t. \quad h_1(\mathbf{x}) = \|\mathbf{x}\|^2 - 1 = 0 \quad h_2(\mathbf{x}) = x + y + z = 0$$

The absolute minimum is 1, 4, 2. The starting point has to verify both constraints are verified : $[0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}]$. The dynamical system follows both constraints. Each Lagrange parameter acts as a force making the dynamical system to stay on the corresponding constraint. The final point is the double projection of the absolute minimum, figure 6

Example III.7.: Consider the minimization on a torus

$$\min x \quad s.t. \quad h(\mathbf{x}) = (R - \sqrt{x^2 + y^2})^2 + z^2 - r^2 = 0$$

The minimum on the torus is on one of the sides. On figure 7, the system evolves on the constraint even if the shape of the surface is complicated. R corresponds to the outer radius and r corresponds to the inner radius.

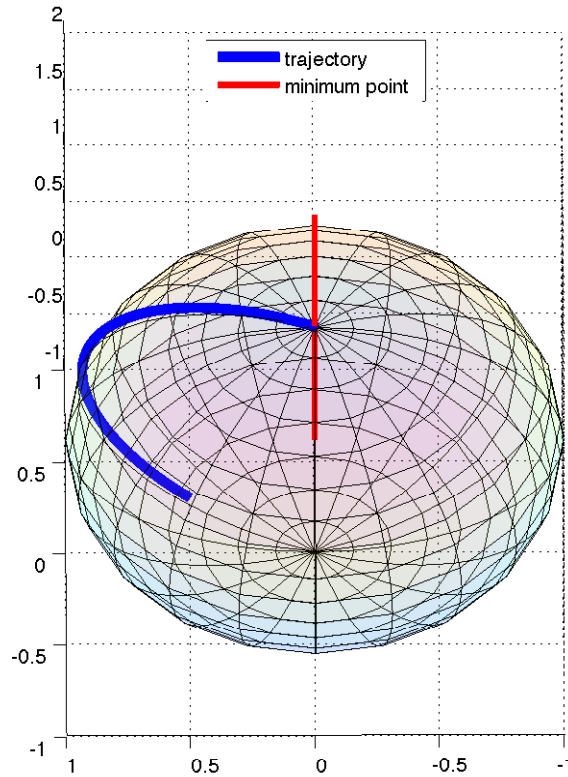


Fig. 4. Minimum distance from a point to a sphere

B. Starting outside the constraints: sliding behavior

1) *The problem and intuition of the solution:* The biggest limitation of this problem is to have to start on the constraints. One solution is to add the following dynamic.

$$-\sum_i c_i h_i(\mathbf{x}) \nabla h_i(\mathbf{x})$$

c_i is a constant that gives the speed of approach to the constraint i . $\nabla h_i(\mathbf{x})$ gives the direction that is normal to the surface locally and $-h_i(\mathbf{x})$ gives a sense to this direction, figure 8. A priori, this property is local because far away from the surface the term in ∇h_i is not necessarily normal to the surface. Consider the modified dynamic system:

$$\dot{\mathbf{x}} = f(\mathbf{x}, t) + \lambda \nabla h(\mathbf{x}) - \sum_j c_j h_j(\mathbf{x}) \nabla h_j(\mathbf{x})$$

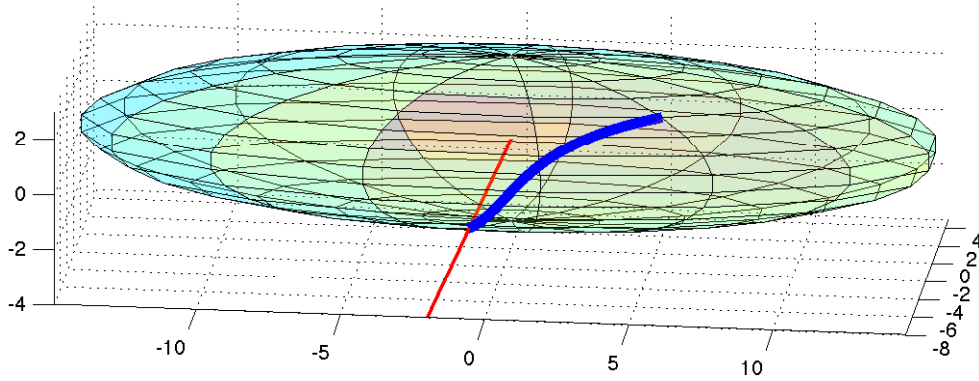


Fig. 5. Dynamical system on the ellipsoid, starting point is 6.7, -2.5, 2.2

In an autonomous gradient system,

$$\dot{\mathbf{x}} = -\nabla(U(\mathbf{x}) + \lambda h(\mathbf{x}) + \frac{1}{2} \sum_j c_j h_j^2(\mathbf{x})) = -\nabla L_c$$

This term corresponds to the term that is added in optimization to construct the augmented lagrangian, L_c . It can be understood as adding a cost when the constraints are not verified.

2) *Sliding surface behavior:* In this section we prove that converging to the surface when starting away from is a global property. We use sliding control techniques. We define $s = \frac{1}{2}(\sum_i c_i h_i^2(\mathbf{x}))$ where $c_i > 0$. s is by definition non negative. Its derivative is

$$\frac{ds}{dt} = \sum_i c_i h_i(\mathbf{x}) \nabla h_i(\mathbf{x}) \dot{\mathbf{x}}$$

Because $f(\mathbf{x}, t)P(\mathbf{x}) \in M(\mathbf{x})$, $\nabla h_i(\mathbf{x})f(\mathbf{x}, t)P(\mathbf{x}) = 0$. The derivative is equal to :

$$\frac{ds}{dt} = -\sum_i c_i h_i(\mathbf{x}) \nabla h_i(\mathbf{x}) \left(\sum_j c_j h_j(\mathbf{x}) \nabla h_j(\mathbf{x}) \right) = -\left(\sum_i c_i h_i \nabla h_i \right)^2 \leq 0$$

If \ddot{s} is bounded, we apply Barbalat's lemma [13]. $\sum c_i h_i \nabla h_i$ converges to zero as time goes to infinity. Using the initial hypothesis, ∇h_i are linearly independent, $c_i h_i$ goes to zero. The constraints are verified. The technique used here is a very common technique in nonlinear control theory. h^2 can be seen as the distance to the surface, which is the same as the sliding variable s used in nonlinear control.

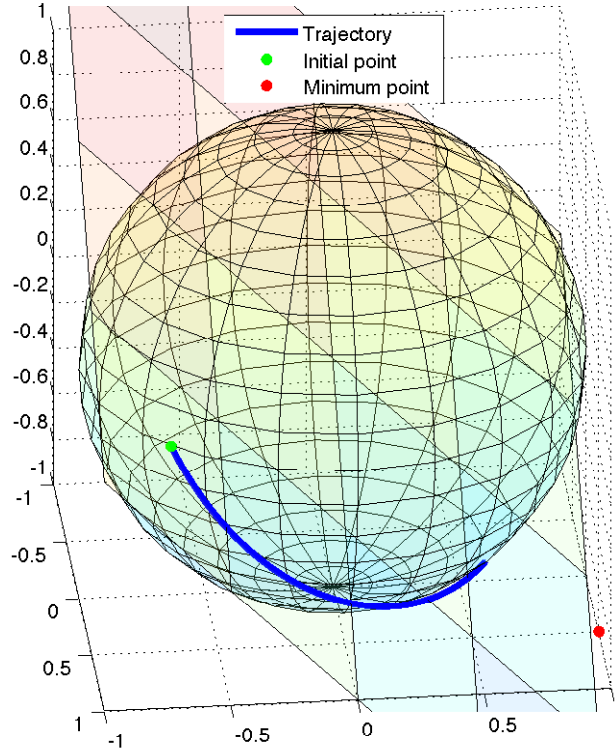


Fig. 6. Example of minimization with two constraints

With one constraint we get exponential convergence towards the surface.

$$\frac{d}{dt} \frac{h^2}{2} = h(\mathbf{x}) \nabla h(\mathbf{x}) \dot{\mathbf{x}} = -c h^2(\mathbf{x}) (\nabla h(\mathbf{x}))^2$$

There are two dynamics in this augmented system. These two dynamics can be controlled using the c parameter. The first dynamic is the so called, Lagrangian dynamics. This dynamic leads to the minimum once on the surface. The second dynamic is the so called sliding dynamics. It makes the system converge to the constraints. If the value of c is very big, we give more importance to the sliding term. If c is small, it will take longer to converge.

3) Examples:

Example III.8.: Consider the optimization problem

$$\min x + y \text{ s.t. } x^2 + y^2 = 1$$

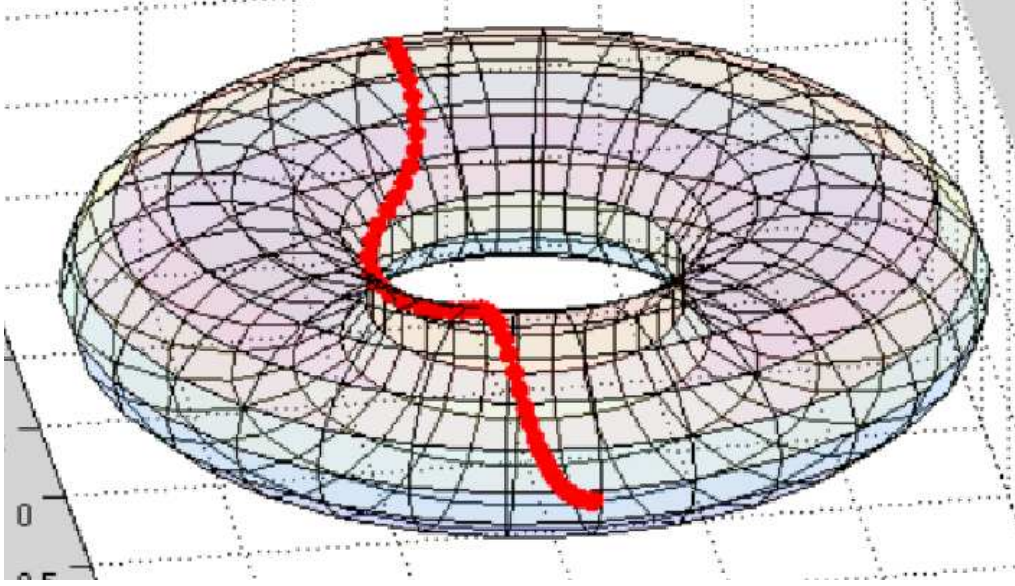


Fig. 7. Minimization on a torus

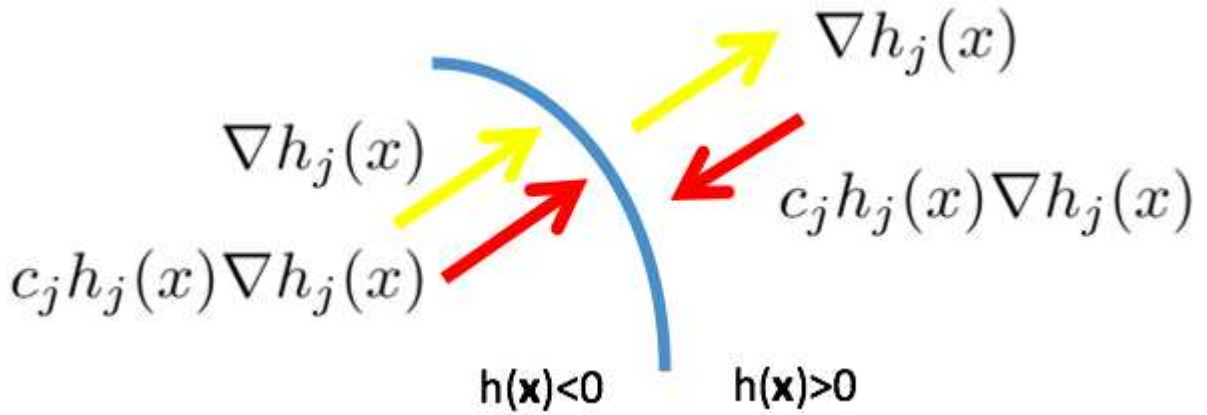


Fig. 8. Sliding dynamic

The projected system is:

$$\dot{x} = \begin{pmatrix} -1 + \frac{x^2 + xy}{x^2 + y^2} - c(x^2 + y^2 - 1)x \\ -1 + \frac{y^2 + xy}{x^2 + y^2} - c(x^2 + y^2 - 1)y \end{pmatrix}$$

If c is small, (figure 9) initially the system follows the Lagrangian dynamic, (it follows a circle that is situated at the starting position). As the system goes away from the surface, the sliding dynamic becomes more important. If c is big, the system converges very quickly to the surface. Once on the surface the Lagrangian dynamic makes the system to evolve on the constraint towards the minimum. The best trade off is to choose a middle value

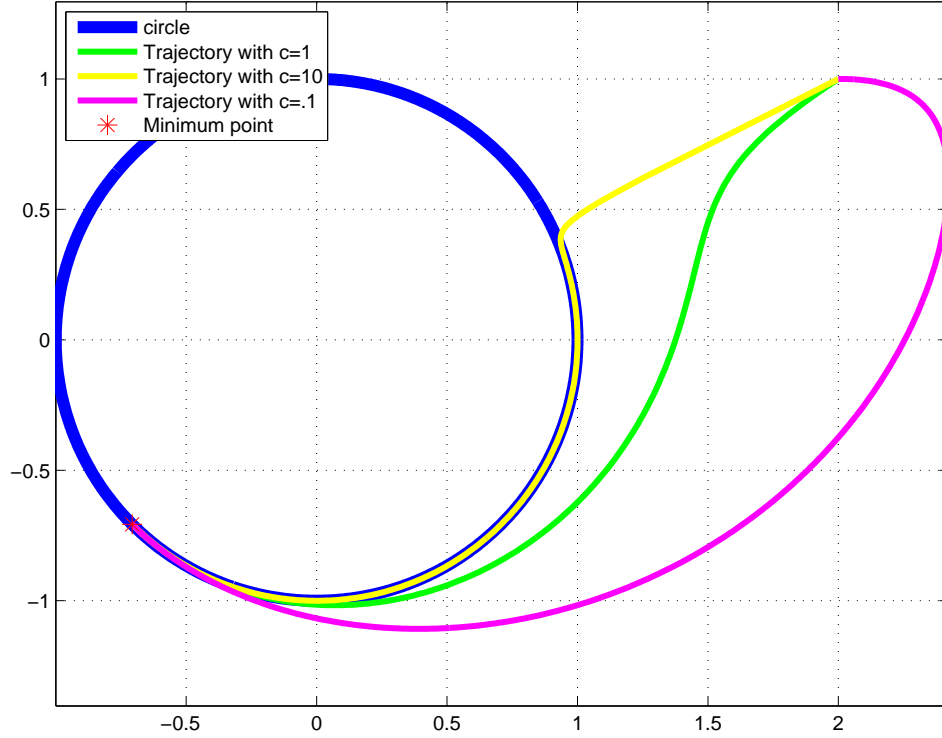


Fig. 9. Minimization of the length of a square in a circle

($c=1$). We approach the surface following the shape of the constraint. The number $\frac{1}{c}$ can be seen as the radius at which the lagrangian dynamics starts working.

Example III.9.: Minimum distance from a point to a sphere Consider the minimization problem, distance from a point to a sphere,

$$\min \frac{1}{2}(x^2 + y^2 + (z - 2)^2) \quad s.t. \quad h(x) = x^2 + y^2 + z^2 - 1 = 0$$

The solution is $(0, 0, 1)$. The projected system adding the sliding term is:

$$\dot{x} = \nabla U + \lambda \nabla h = \begin{pmatrix} x + \lambda x + c(x^2 + y^2 + z^2 - 1)x \\ y + \lambda y + c(x^2 + y^2 + z^2 - 1)y \\ z - 2 + \lambda z + c(x^2 + y^2 + z^2 - 1)z \end{pmatrix}$$

with $\lambda = \frac{2z-1}{x^2+y^2+z^2}$.

Example III.10.: The problem is to maximize the volume of a parrallelepiped staying on an ellipse

$$\min -xyz \quad s.t. \quad h(x) = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1 = 0$$

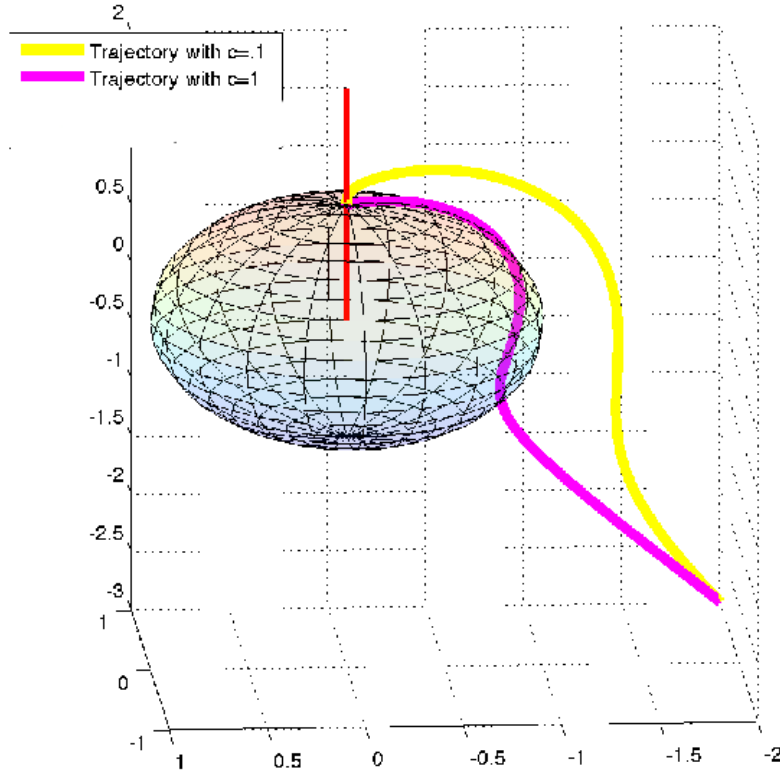


Fig. 10. Minimum distance from a point to a sphere

Consider $a = 15; b = 5; c = 3$; figure 11 shows the convergence.

Example III.11.: The minimization problem with two constraints

$$\min \frac{1}{2} \|\mathbf{x} - \mathbf{x}_1\|^2 \quad s.t. \quad h_1(\mathbf{x}) = \|\mathbf{x}\|^2 - 1 = 0 \quad h_2(\mathbf{x}) = x + y + z = 0$$

The final point is the double projection of the absolute minimum. The values c_1, c_2 are the gains to the two surfaces. To stay on the sphere, we increase the gain ($c_1 = 10$).

IV. CONTRACTION THEORY WITH TIME-VARYING EQUALITY CONSTRAINTS

This section has two main parts. In the first one, we state a theorem giving the condition of contraction for a time varying constrained nonlinear dynamic system. We start on the constraints to have contraction behavior. In the second part, we allow the system to start outside the constraints. We use sliding techniques to conclude convergence to the constraints.

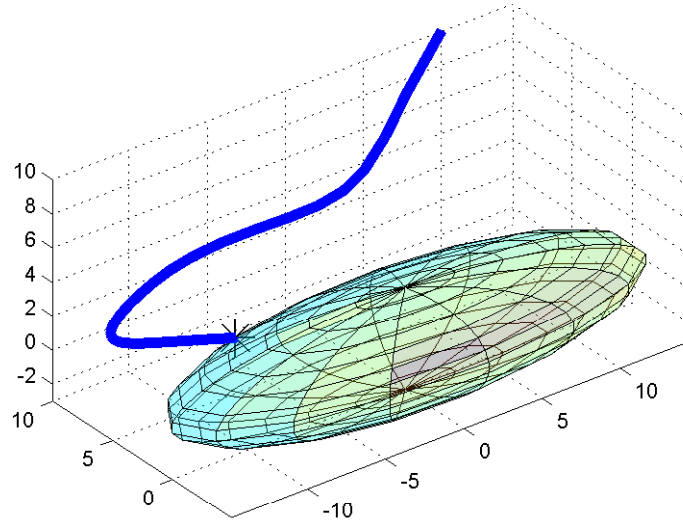


Fig. 11. Maximum volume on an ellipse

A. Starting on the constraints: contraction theory

We define the set $S(\mathbf{x}, t) = [\mathbf{x} \mid \mathbf{h}(\mathbf{x}, t) = 0]$.

1) *Contraction theorem:* In this section we compute the condition of contraction for dynamic systems under time-varying constraints.

Theorem 5. *The condition of contraction for the system*

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) \text{ s.t. } \mathbf{h}(\mathbf{x}, t) = 0$$

is that there exists a uniform positive definite metric; Θ and that the hermitian part of :

$$\mathbf{F} = \left(\dot{\Theta}(\mathbf{x}, t) + \Theta(\mathbf{x}, t) P(\mathbf{x}) \left(\frac{\partial \mathbf{f}}{\partial \mathbf{x}} + \lambda(\mathbf{x}, t) \frac{\partial^2 \mathbf{h}(\mathbf{x})}{\partial \mathbf{x}^2} \right) \right) \Theta(\mathbf{x}, t)^{-1} \text{ on } S(\mathbf{x}, t)$$

has a maximum eigenvalue uniformly negative. $P(\mathbf{x})$ is the projection operator onto the tangent space $S(\mathbf{x}, t)$ and $\lambda(\mathbf{x}, t) = -(\mathbf{f}(\mathbf{x}, t) \nabla \mathbf{h} + \frac{\partial \mathbf{h}}{\partial t}) [\nabla \mathbf{h} \nabla \mathbf{h}']^{-1}$. The initial condition must verify the constraint.

The proof has three parts. The initial problem is

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) \text{ s.t. } h(\mathbf{x}, t) = 0$$

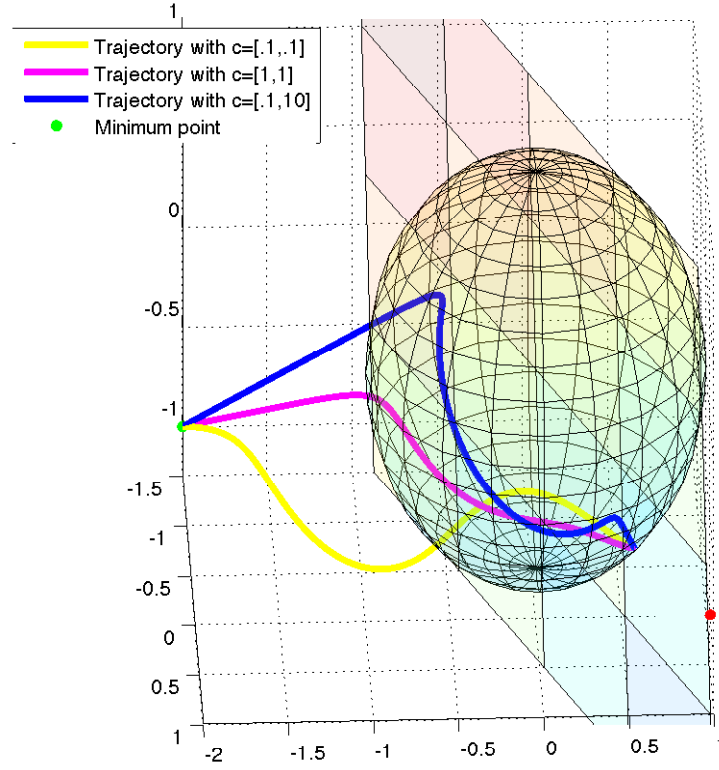


Fig. 12. Example of minimization with two constraints

Physically to remain on the constraints, we need to project the speed and consider the surface's speed, figure 13 This last term is calculated because the constraints are always verified, $0 = \frac{dh}{dt} = \nabla h \dot{x}' + \frac{\partial h}{\partial t}$ As we have the choice on \dot{x}' , we add a term to ensure that this equation is verified.

$$-\frac{\partial \mathbf{h}}{\partial t} [\nabla \mathbf{h} \nabla \mathbf{h}']^{-1} \nabla \mathbf{h} = \lambda_t \nabla \mathbf{h}$$

We can see this new parameter as a force that makes the system stay on the constraint. It is a new generalization of Lagrange parameters. The total system is:

$$\dot{x} = \mathbf{f}(\mathbf{x}, t) + \lambda_1 \nabla \mathbf{h} - \frac{\partial \mathbf{h}}{\partial t} [\nabla \mathbf{h} \nabla \mathbf{h}']^{-1} \nabla \mathbf{h} = \mathbf{f}(\mathbf{x}, t) + \lambda_1 \nabla \mathbf{h} + \lambda_t \nabla \mathbf{h} = \mathbf{f}(\mathbf{x}, t) + \lambda \nabla \mathbf{h}$$

where $\lambda = \lambda_1 + \lambda_t$. The constraints are verified at anytime because the system starts on the constraint and the velocity remains on the constraints.

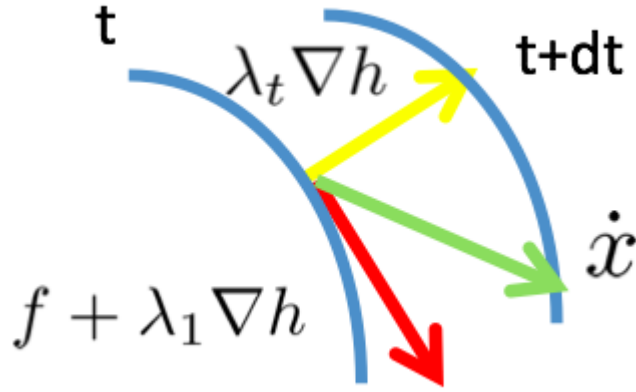


Fig. 13. Projection with time-varying constraints

We compute $\delta \dot{\mathbf{x}} = \left(\frac{\partial f}{\partial \mathbf{x}} + \lambda(\mathbf{x}) \frac{\partial^2 h}{\partial \mathbf{x}^2} + \frac{\partial \lambda_1(\mathbf{x})}{\partial \mathbf{x}} \frac{\partial h}{\partial \mathbf{x}} + \frac{\partial \lambda_t(\mathbf{x})}{\partial \mathbf{x}} \frac{\partial h}{\partial \mathbf{x}} \right) \delta \mathbf{x}$. $\frac{\partial \lambda_1(\mathbf{x})}{\partial \mathbf{x}}$ can be computed as well as $\frac{\partial \lambda_t(\mathbf{x})}{\partial \mathbf{x}}$.

We can substitute in the precedent equation and reorganize:

$$\delta \dot{\mathbf{x}} = P(\mathbf{x}) \left(\frac{\partial f}{\partial \mathbf{x}} + \lambda(\mathbf{x}) \frac{\partial^2 h}{\partial \mathbf{x}^2} \right) \delta \mathbf{x} - \nabla h(\mathbf{x})' [\nabla h(\mathbf{x}) \nabla h(\mathbf{x})']^{-1} \left(\delta \mathbf{x} \frac{\partial^2 h}{\partial \mathbf{x}^2} \dot{\mathbf{x}} + \frac{\partial h}{\partial t} \right)$$

It is essential to note that the first term belongs to $M(\mathbf{x})$ and the second term belongs to $M(\mathbf{x})^\perp$.

To have contraction behavior the first term of $\delta \dot{\mathbf{x}}^\parallel$ must be uniformly bounded.

$$\delta \dot{\mathbf{x}} = P(\mathbf{x}) \left(\frac{\partial f}{\partial \mathbf{x}} + \lambda(\mathbf{x}) \frac{\partial^2 h}{\partial \mathbf{x}^2} \right) \delta \mathbf{x}$$

Finally

$$\frac{d}{dt}(\delta \mathbf{z}' \delta \mathbf{z}) = 2\delta \mathbf{z}' \left(\dot{\Theta}(\mathbf{x}, t) + \Theta(\mathbf{x}, t) P(\mathbf{x}) \left(\frac{\partial f}{\partial \mathbf{x}} + \lambda(\mathbf{x}) \frac{\partial^2 h}{\partial \mathbf{x}^2} \right) \right) \Theta(\mathbf{x}, t)^{-1} \delta \mathbf{z}$$

2) *Theorem's applications:* One of the most important application is optimization under varying constraints.

Example IV.1.: The problem of the sum of the sides under a growing circle is :

$$\min x + y \text{ s.t. } x^2 + y^2 = t^2$$

The solution using KKT theorem, 3, is given by $x^* = y^* = -\frac{t}{\sqrt{2}}$. Using the precedent theorem we create the 'extended' lagrangian system, simulated on figure 14

$$\dot{\mathbf{x}} = - \begin{pmatrix} 1 - \frac{(x+y)}{t^2} x - \frac{x}{t} \\ 1 - \frac{(x+y)}{t^2} y - \frac{y}{t} \end{pmatrix}$$

The condition of contraction is $\nabla^2 f + (\lambda_1 + \lambda_t)\nabla^2 h = -\frac{x+y+t}{t^2}I > 0$. The final condition is $x + y + t < 0$.

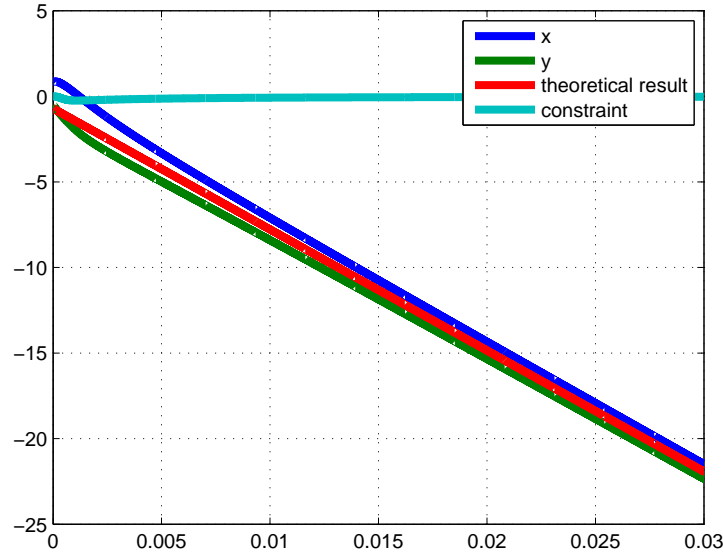


Fig. 14. Minimization of the sum of the sides under a growing circle

Example IV.2.: The problem of the sum of the sides under a growing ellipse is :

$$\min x + y \text{ s.t. } tx^2 + y^2 = 1$$

The solution using KKT theorem, 3, is given by $x^* = -\frac{1}{\sqrt{1+t^2}}$, $y^* = -\frac{t}{\sqrt{1+t^2}}$.

$$\dot{x} = - \begin{pmatrix} 1 - \frac{(tx+y)}{t^2x^2+y^2}tx - \frac{x^2}{2t^2x^2+2y^2}xt \\ 1 - \frac{(tx+y)}{t^2x^2+y^2}y - \frac{x^2}{2t^2x^2+2y^2}y \end{pmatrix}$$

The condition of contraction is given by $\nabla^2 f + (\lambda_1 + \lambda_t)\nabla^2 h = -\frac{x^2+tx+y}{t^2x^2+y^2}I > 0$. The final condition is $x^2 + tx + y < 0$.

In this example; we use time varying constraints to solve problems with static constraints.

Example IV.3.: In order to make a smoother approach to a certain constraint h_0 , we can create a time varying constraint such that when $t \rightarrow \infty$, $h(x, t) \rightarrow h_0(x)$. Consider the minimization problem:

$$\min x + y \text{ s.t. } x^2 + y^2 = e^{\frac{2}{t}}$$

The dynamical system associated to that is :

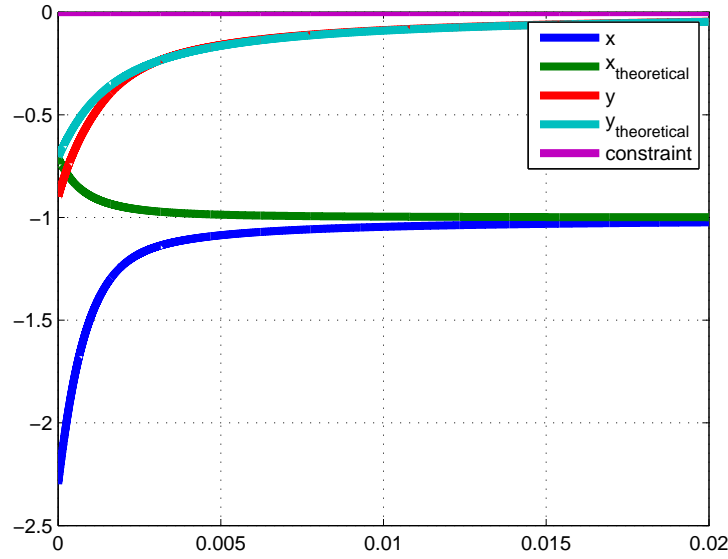


Fig. 15. Minimization of the sum of the sides under a growing ellipse

$$\dot{x} = - \begin{pmatrix} 1 - \frac{(x+y)}{e^{\frac{2}{t}}}x - \frac{x}{t^2} \\ 1 - \frac{(x+y)}{e^{\frac{2}{t}}}y - \frac{y}{t^2} \end{pmatrix}$$

B. Starting outside the constraints: sliding behavior

1) *Sliding surface behavior:* As done in the precedent chapter, we add a sliding term. The new system is :

$$\dot{x} = f(x, t) + \lambda \nabla h - \frac{\partial h}{\partial t} [\nabla h \nabla h']^{-1} \nabla h - \sum_j c_j h_j(x) \nabla h_j(x)$$

We have global convergence towards the surface. We define $s = \frac{1}{2}(\sum_i c_i h_i^2(x))$ where $c_i > 0$. s is by definition strictly positive. Its derivative:

$$\begin{aligned} \frac{ds}{dt} &= C' h' \left(\frac{\partial h}{\partial x} \dot{x} + \frac{\partial h}{\partial t} \right) \\ &= C' h' \left(\frac{\partial h}{\partial x} \left(P(x)' f(x, t)' - \nabla h' [\nabla h \nabla h']^{-1} \frac{\partial h}{\partial t} - \sum_j c_j h_j(x) \nabla h_j(x) \right) + \frac{\partial h}{\partial t} \right) \end{aligned}$$

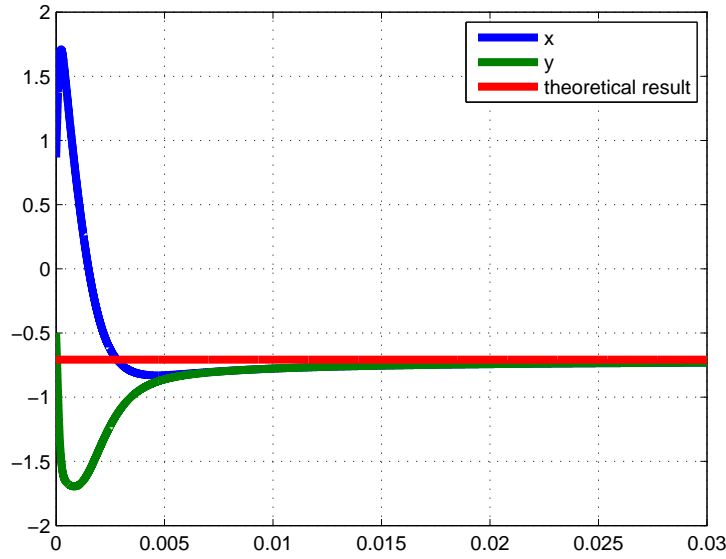


Fig. 16. Using time varying constraints to solve problems with static constraints

Because $\mathbf{f}(x, t)P(x) \in M(x)$, $\nabla h_i(x)\mathbf{f}(x, t)P(x) = 0$. As $\nabla \mathbf{h} \nabla \mathbf{h}' [\nabla \mathbf{h} \nabla \mathbf{h}']^{-1} \frac{\partial \mathbf{h}}{\partial t} = \frac{\partial \mathbf{h}}{\partial t}$. The derivative is

$$\frac{ds}{dt} = -(\sum c_i h_i \nabla h_i)^2 \leq 0$$

The second derivative is smooth, it is bounded. Applying Barbalat's lemma, we conclude that $\sum c_i h_i \nabla h_i$ converges to zero as time goes to infinity. As ∇h_i are linearly independent, $c_i h_i$ goes to zero. The constraints are verified.

2) *Examples:* To show how this sliding term works, we use the precedent examples but starting on a point outside the surface.

Example IV.4.: Minimization of the sum of the sides under a growing circle The problem is :

$$\min x + y \text{ s.t. } x^2 + y^2 = t^2$$

The new system is

$$\dot{x} = - \begin{pmatrix} 1 - \frac{(x+y)}{x^2+y^2}x - \frac{x}{t} + chx \\ 1 - \frac{(x+y)}{x^2+y^2}y - \frac{y}{t} + chy \end{pmatrix}$$

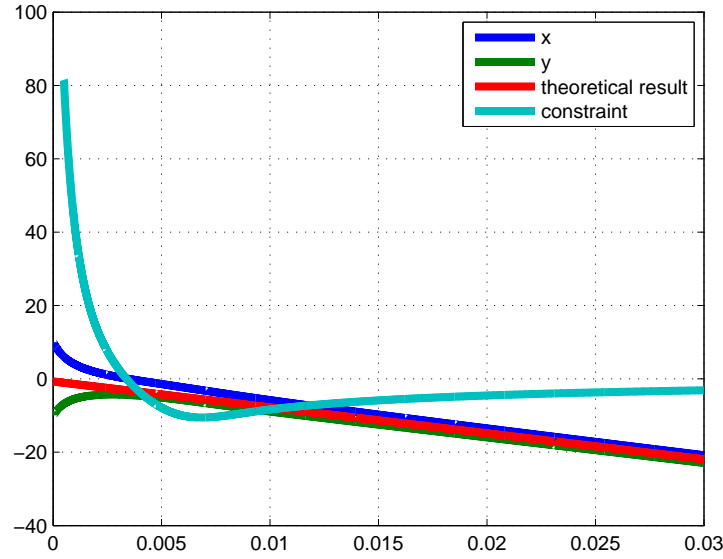


Fig. 17. Minimization of the sum of the sides under a growing circle

Example IV.5.: Minimization of the sum of the sides under a growing ellipse The problem is :

$$\min x + y \text{ s.t. } tx^2 + y^2 = 1$$

The new system is

$$\dot{x} = - \begin{pmatrix} 1 - \frac{(tx+y)}{t^2x^2+y^2}tx - \frac{x^2}{2t^2x^2+2y^2}xt + chxt \\ 1 - \frac{(tx+y)}{t^2x^2+y^2}y - \frac{x^2}{2t^2x^2+2y^2}y + chy \end{pmatrix}$$

V. CONCLUSION AND DIRECTIONS FOR FUTURE RESEARCH

This research has four main new contributions.

- The first and more important is the condition of contraction (a condition of convergence) for two different dynamic systems, constrained with equalities and constrained with time varying inequalities.
- The second contribution is, in the particular case of a gradient autonomous contracting in a metric identity system, an algorithm to find the minimum. A more general metric could allow solving more complicated problems.
- The third contribution is the understanding that adding a term $\frac{h^2}{2}$ in the cost function corresponds to dynamic term that makes the dynamic system converge towards the surface.

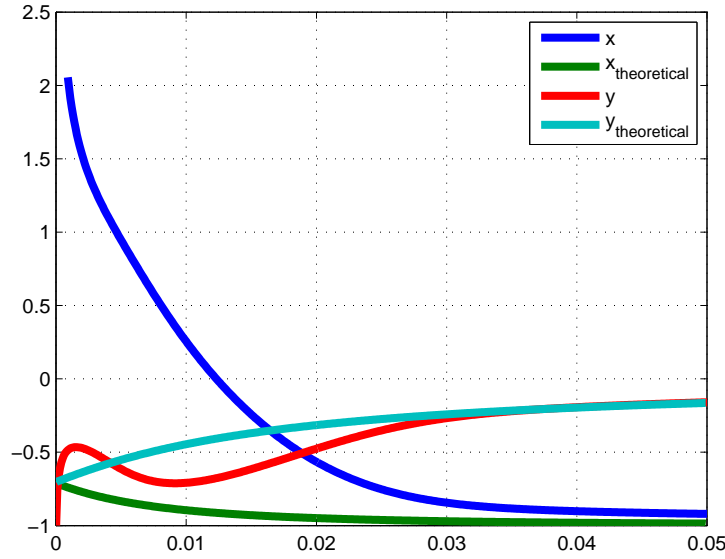


Fig. 18. Minimization of the sum of the sides under a growing ellipse

- The fourth and last contribution is on Lagrange parameters. This paper gives a new physical approach of Lagrange parameters. They can be understood as a scalar value that ensure the system to have a tangential speed to the constraints. Also it gives three generalizations of Lagrange parameters, with a time varying cost function $\lambda = -\nabla U(\mathbf{x}, t) \nabla h' [\nabla h \nabla h']^{-1}$, with a time varying general vector $\lambda = -\mathbf{f}(\mathbf{x}, t) \nabla h' [\nabla h \nabla h']^{-1}$ and with a a time varying general vector with time varying constraints $\lambda = -\left(\mathbf{f}(\mathbf{x}, t) \nabla h' + \frac{\partial h}{\partial \mathbf{x}}\right) [\nabla h \nabla h']^{-1}$. This last case is the most general because it contains the other two.

There are many directions for the future research.

- We have always used an identity metric. It is interesting to explore what happens with constant metrics and time varying metrics. In the gradient autonomous case, it could include some information of the constraints making the minimization faster.
- When having many different goals, Pareto optimality does not have a general theorem to find the best solution. May be some conclusive theorem can be found exploring Pareto optimality using contraction theory. Contraction theory has very interesting combination properties.
- In chapter one, an example using parallel combination has been presented. We could try to

use feedback and hierarchical combination.

- The Hamilton-Jacobi-Bellman (HJB) equation is a partial differential equation which is central to optimal control theory. It is the solution of a minimization problem subject to a dynamic system.
- In the article [20], the ordinary gradient of a function does not represent its steepest direction, but the natural gradient does. Information geometry is used for calculating the natural gradients. How does this natural gradient relates to our projection operator ?.

REFERENCES

- [1] Lohmiller, W., and Slotine, J.J.E., On Contraction Analysis for Nonlinear Systems, *Automatica*, 34(6), 1998.
- [2] Bertsekas, Dimitri P., Athena Scientific, Nonlinear programming, 1995.
- [3] J.J. Slotine, Modular Stability Tools for Distributed Computation and Control J. Adaptive Control and Signal Processing, 2002.
- [4] David G. Luenberger, The Gradient Projection Method Along Geodesics, *Management science*, 1972.
- [5] Personal Communication, Eric Feron.
- [6] Gignoux, Claude, Silvestre-Brac, Bernard, Solved Problems in Lagrangian and Hamiltonian Mechanics, Grenoble Sciences, 2004.
- [7] Wolfgang Kuhnel, Differential Geometry: Curves - Surfaces - Manifolds, AMS, 2006.
- [8] Soto Jonathan, NonLinear contraction tools for constrained optimization, S.M Thesis, MIT, 2010.
- [9] P.-A. Absil, K. Kurdyka, On the stable equilibrium points of gradient systems , *Systems and control letters*. 2006.
- [10] Lohmiller, W., and Slotine, J.J.E. (2000b). Nonlinear Process Control Using Contraction Theory, *A.I.Ch.E. Journal*.
- [11] Nadav Kashtan, Elad Noor, and Uri Alon, Varying environments can speed up evolution, *PNAS*, 2007.
- [12] Constantin Udriste, Convex Functions And Optimization Methods On Riemannian Manifolds, Springer, 1994.
- [13] Slotine, Jean-Jacques and Li, Weiping, Prentice Hall, Applied Nonlinear Control, 1990 .
- [14] Jacob Mattingley and Stephen Boyd, Real-Time Convex Optimization in Signal Processing, *IEE Signal Processing magazine*, May 2010.
- [15] Luenberger, D.G. (1979). Introduction to dynamic systems, Wiley.
- [16] Arnold, V.I. (1978). Mathematical Methods of Classical Mechanics, Springer Verlag.
- [17] Zuo, L., and Slotine, J.J.E. (2003). Robust vibration isolation using frequency shaped sliding variables, MIT Nonlinear Systems Laboratory, Report MIT-NSL-030201, submitted.
- [18] Petros A. Ioannou, Bari Fidan, Adaptive control tutorial, SIAM, 2006.
- [19] Lohmiller Winfried, J.J. Slotine, Exact Modal Decomposition of Nonlinear Hamiltonian Systems Hamiltonian decomposition, American Institute of Aeronautics and Astronautics.
- [20] Amari S.I. Natural Gradient Works Efficiently in Learning, *RIKEN Frontier Research Program*, Saitama 351-01, Japan.
- [21] Goldstein H., Classical Mechanics, Addison Wesley, 1980.
- [22] Uwe Helmke (Author), John B. Moore (Author), R. Brockett, Optimization and Dynamical Systems, Communications and Control Engineering, 1996.
- [23] On the duality of strong convexity and strong smoothness: Learning applications and matrix regularization, Sham M. Kakade and Shai Shalev-Shwartz and Ambuj Tewari, Toyota Technological Institute Chicago, USA
- [24] P. Hartman, *Canadian Journal of Mathematics*, 13, 480 (1961).
- [25] D. C. Lewis, *American Journal of Mathematics*, 71, 294 (1949).
- [26] D. Angeli, *IEEE Transactions on Automatic Control*, 47, 410 (2002).
- [27] A. Pavlov, A. Pogromvsky, N. van de Wouf, and H. Nijmeijer, *Systems and Control Letters*, 52, 257 (2004).
- [28] W. Lohmiller and J. J. Slotine, *International Journal of Control*, 78, 678 (2005).
- [29] W. Wang and J. J. E. Slotine, *Biological Cybernetics*, 92, 38 (2005).